基于高斯混合-隐马尔科夫融合算法识别奶牛步态时相

张　楷，韩书庆，程国栋，吴赛赛，刘继芳*

（中国农业科学院农业信息研究所/农业农村部区块链农业应用重点实验室，北京 100081）

摘　要：奶牛步态时相是反映奶牛健康及跛行严重程度的重要指标。为准确自动识别奶牛步态时相，本研究提出一种融合高斯混合模型（Gaussian Mixture Model，GMM）和隐马尔科夫模型（Hidden Markov Model，HMM）的无监督学习奶牛步态时相识别算法 GMM-HMM。使用惯性测量单元采集奶牛后肢加速度和角速度信号，通过卡尔曼滤波消除噪声，筛选并提取特征值，构建GMM-HMM模型，实现奶牛静立相、连续步态中的站立相和摆动相等3种步态时相的自动识别。结果表明，静立相识别的准确率、召回率和F_1分别为89.28%、90.95%和90.91%，连续步态中的站立相识别的准确率、召回率和F_1分别为91.55%、86.71%和89.06%，连续步态中的摆动相识别的准确率、召回率和F_1分别为86.67%、91.51%和89.03%。奶牛步态分割的准确率为91.67%，相较于基于事件的峰值检测法和动态时间规整算法准确率分别提高了4.23%和1.1%。本研究可为下一步基于穿戴式步态分析的奶牛跛行特征提取提供技术参考。

关键词：奶牛跛行；步态时相；步态分割；高斯混合模型；隐马尔科夫模型；卡尔曼滤波

中图分类号：S24;TP274　　**文献标志码：**A　　**文章编号：**SA202204003

引用格式：张楷，韩书庆，程国栋，吴赛赛，刘继芳．基于高斯混合-隐马尔科夫融合算法识别奶牛步态时相[J]．智慧农业(中英文)，2022，4(2)：53-63.

ZHANG Kai, HAN Shuqing, CHENG Guodong, WU Saisai, LIU Jifang. Gait phase recognition of dairy cows based on Gaussian Mixture model and Hidden Markov model[J]. Smart Agriculture, 2022, 4(2): 53-63. (in Chinese with English abstract)

1　引　言

奶牛跛行是由于蹄底溃疡、趾间皮炎和腐蹄病[1]等各种肢蹄病引起奶牛不适，导致其承重能力不足以及步态异常。奶牛跛行与饲养方式、牧场环境卫生等有关[2]，影响奶牛正常生产生活，进而导致产奶量下降、个体产奶年限缩短以及治疗成本上升等问题，严重影响牧场经济效益。当前中国奶牛肢蹄病发病率较高，跛行率超过了30%[3]，人工识别存在效率低、主观判断存在偏差、劳动力成本较高和容易造成奶牛应激反应等问题，发展奶牛跛行自动识别技术需求强烈[4]。

奶牛跛行步态分析的关键是在奶牛连续步态中识别摆动相和站立相等步态时相，通过分析奶牛各个步态时相的运动强度和持续时间判断奶牛跛行程度。奶牛的步态满足一定的周期性和规律

收稿日期：2022-04-08

基金项目：国家自然科学基金项目（32102600）；中国农业科学院科技创新工程项目（CAAS-ASTIP-2016-AII）；中央级公益性科研院所基本科研业务费专项（JBYW-AII-2021-33，JBYW-AII-2022-36）

作者简介：张　楷（1997—），男，硕士研究生，研究方向为畜禽行为识别与畜牧物联网技术。E-mail：82101212474@caas.cn

*通信作者：刘继芳（1965—），男，博士，研究员，研究方向为农业信息管理。E-mail：liujifang@caas.cn

性，但在现实环境中，奶牛步态受环境和个体差异的影响较大，自适应奶牛步态分析需要针对现实情况采取相应的技术方法。现较多采用计算机视觉技术和可穿戴式设备[5]来进行奶牛步态分析。计算机视觉技术具有非接触、免应激、低成本、高通量、不干扰奶牛正常活动等优点[6]，已被用于奶牛步态分割[7]、步态特征提取[8]以及跛行检测[9]中。但计算机视觉容易受到环境光线、复杂背景以及奶牛相互遮挡的影响[10]。可穿戴式传感器[11]具有不限于特定测量区域以及不受光照和遮挡影响的优点，目前，基于可穿戴式传感器进行奶牛跛行的研究热点集中在奶牛步态分割和特征提取方面。常见的奶牛步态分割方式主要有基于事件的峰值检测法和基于时域的动态时间规整方法两种。峰值检测法最先应用于奶牛步态分割和识别[12]，但其容易将一些非步态的峰值识别为步态事件，导致低准确率和高召回率。动态时间规整法是指通过计算两个序列的相似性进行步态分割，设定阈值以及建立最优匹配模板，自动化程度不高[13]。

隐马尔科夫模型（Hidden Markov Model，HMM）是马尔科夫模型发展而来的机器学习算法，核心思想是通过研究观测序列来辨识不可见状态序列。近年来，许多研究在人的步态识别上使用HMM。在人体步态的研究中分析发现，步态信号符合马尔科夫链的时序变化，针对时序数据的动态过程具有鲁棒性以及能够处理牧场条件下奶牛步态数据中的异质性，可以进行步态阶段识别[14]。不同步态分割方法对比显示，基于HMM的分割方法优于动态时间规整模板匹配法以及峰值检测法[15]。由于本研究解决的是长时间连续序列的奶牛步态时相识别问题，观测数据量较大，所以引入高斯混合模型（Gaussian Mixture Model，GMM）。GMM可以无监督地对观测数据进行聚类，GMM和HMM混合算法曾用于人类步态识别，采用无监督的学习方式，得到帕金森患者步态分割，准确率为96.2%[16]。目前，基于GMM和HMM融合的奶牛步态时相识别研究尚未展开。

综上所述，现有技术方法初步实现了奶牛步态分割，但还未深入到奶牛步态时相识别，并且依旧存在数据标注困难、准确率均未达到实际应用要求的问题。为进一步改进奶牛步态时相自动识别方法以及提高其准确率，本研究提出一种融合GMM和HMM的无监督学习奶牛步态时相识别算法GMM-HMM，利用惯性测量单元采集奶牛步态数据，结合视频标注构建奶牛步态时相识别数据集；筛选出冠状轴角速度的一阶差分作为奶牛步态时相识别的特征值，减少奶牛个体差异对识别效果的影响，并分析奶牛步态分割效果，为奶牛步态分析提供一种新的技术手段。

2 材料与方法

2.1 试验设备与场地

试验在中国农业机械化科学研究院北京农机试验站奶牛养殖场进行。试验对象为5头随机选择的荷斯坦奶牛，利用MEMS微型惯性测量单元传感器（LPMS-B2型号）采集奶牛后肢加速度和角速度数据，采样频率为50 Hz，陀螺仪测量范围为±2000 dps。在奶牛活动场内，选定一条长20 m、宽1.5 m走廊作为测量场地，测量场地示意图如图1所示。使用佳能EOS 5D Mark III相机同步拍摄奶牛通过走廊的视频。

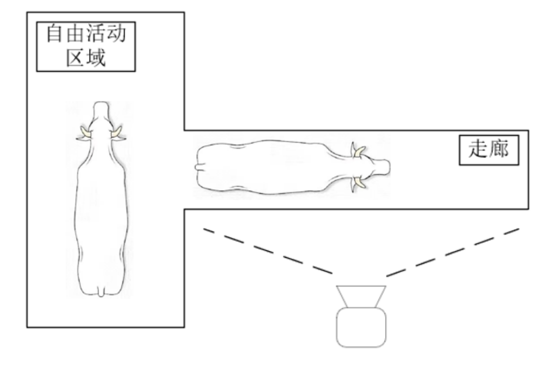


图1 测量场地示意图

Fig. 1 Schematic diagram of measuring site

2.2 数据采集与预处理

2.2.1 数据采集

由于奶牛蹄病发生于后肢的概率较高[17]，以及为方便收集数据，统一将传感器绑在奶牛右后肢上。5头试验奶牛佩戴方向完全相同，在传感器中，定义x轴为奶牛前进方向，y轴为竖直方向；z轴为冠状轴方向，如图2所示。奶牛在测量场地自由活动，相机以50 f/s帧率录制奶牛通过走廊的视频。根据传感器的时间戳与相机记录的视频时间实现图像数据与传感器数据时间同步。试验中，采用人工引导的方式确保所有试验奶牛多次通过走廊，记录充足的图像数据。将人工标注的视频数据作为奶牛步态时相识别的黄金标准。试验采集了2022年2月23日14:00—16:30的奶牛步态数据，每个传感器有240,000～260,000条数据记录，拍摄的奶牛通过走廊的视频共计18段。

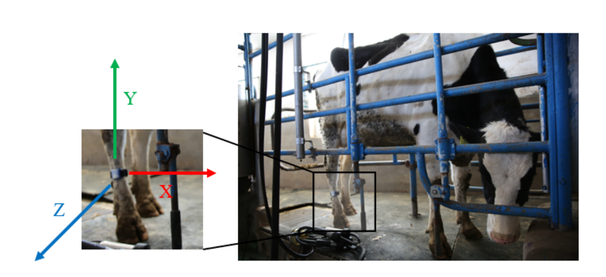


图2 传感器佩戴方式及传感器坐标

Fig. 2 Sensor wearing method and sensor coordinates

2.2.2 数据预处理

LPMS-B2传感器数据中与奶牛步态的运动学数据直接相关的是加速度和角速度，使用3轴加速度和3轴角速度数据为基础进行奶牛步态时相识别。由于传感器同时测量9轴数据，高频噪声较少，噪声多受系统干扰的影响，为去除系统的噪声和干扰影响，还原真实的动态数据，采用卡尔曼滤波（Kalman filtering）进行数据预处理。本试验使用的传感器集成加速度计、陀螺仪和磁力计三个部分，同时测量x、y、z轴三个方向的线加速度、角速度以及磁场强度。在测量过程中合计9轴的数据互相影响并且单位不同，使用卡尔曼滤波可以减轻传感器噪声，降低系统误差和测量误差。该传感器内置滤波器，在采集数据之前已将滤波器设定完成，得到的数据噪声因素已经消除，可以直接利用数据进行模型分析。将图像数据采用FFmpeg软件进行视频分帧处理，通过逐帧比对的方式，确定蹄离地和蹄落地两个步态事件，再与传感器数据进行匹配。由图2可知，x轴的加速度可以表示奶牛基本的运动学特征，z轴方向为水平且垂直于前进方向向外，z轴角速度可以捕捉奶牛细微的牛蹄运动，可以进一步分析奶牛的步态情况，所以分析奶牛后肢的x轴加速度和z轴角速度可以反应奶牛步态事件，更容易实现奶牛步态事件与传感器数据的同步。x轴加速度和z轴角速度折线图对应的奶牛步态事件如图3所示。

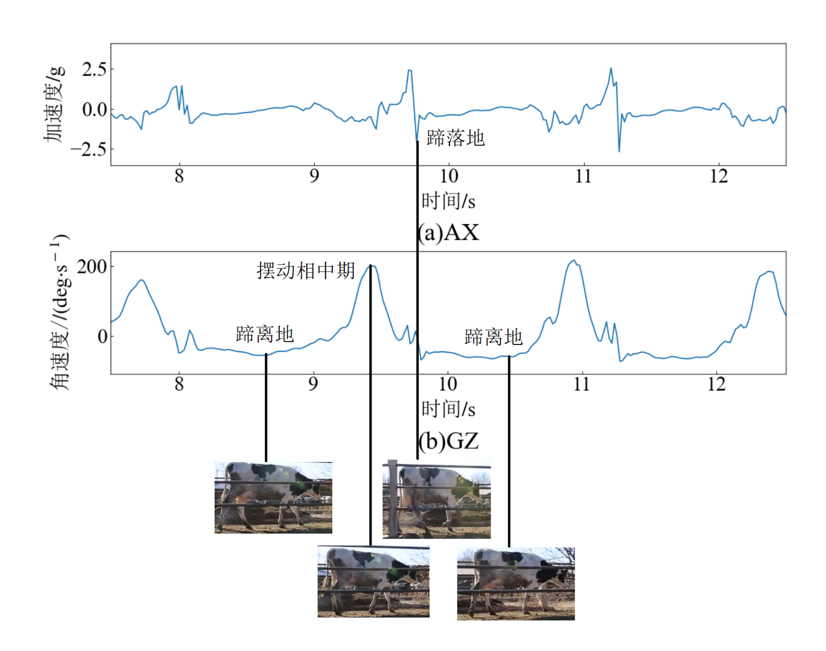
2.3 训练集和测试集数据构建

数据集构建以测试集为基准，测试集构建时使用人工标注的方式，将图像数据与传感器数据进行匹配，在传感器数据中标出步态事件（蹄落地，蹄离地），图像数据是随机采集的，符合测试集构建的规则。观察6轴传感器数据并做了大量尝试后发现，需要找出周期性最明显的数据，方便数据标注也更容易找到特征值，最终确定将z轴的角速度作为本次试验数据，如图4所示。

为消除佩戴传感器时奶牛行为以及刚佩戴好传感器后奶牛不适应带来的试验误差，将原始数据的前80,000个记录点（约26.7 min）的数据作为误差项去除，同理，将后80,000个记录点同时去除。由于传感器存在时间戳，利用视频时间可以实现奶牛步态时间与传感器数据同步，每头牛有80～90 min的传感器数据，提取有视频标注的传感器数据作为测试集，剩余的传感器数据去除误差项作为无监督的训练集，其中包含每头牛至少30 min以上的高质量传感器数据，测试集中连续步态有162步，训练集中连续步态有729步。

2.4 GMM-HMM融合算法描述

本研究采用GMM-HMM融合算法进行奶牛



注:AX为x轴加速度,GZ为z轴角速度

图3　x轴加速度和z轴角速度折线图对应的奶牛步态事件

Fig. 3 Gait events of dairy cows corresponding to the curve of x-axis acceleration and z-axis angular velocity

步态时相识别研究。

奶牛步态时相为奶牛行走的单个步态周期中不同阶段，本研究将连续步态分为摆动相和站立相。摆动相为奶牛蹄离地到蹄着地的过程，站立相为蹄着地到蹄离地的过程。在连续的步态序列中，摆动相和站立相是交替出现的，但是本试验过程中，奶牛除了行走以外，还存在静止站立的状态。因此，本研究将奶牛四肢同时着地静止不动的状态定义为静立相，区别于连续步态序列中，奶牛四肢交换站立的站立相。静立相、摆动相和站立相都可以反映奶牛的运动状态，同时奶牛步态时相的识别可以实现对奶牛的步态分割，识别的连续站立相和摆动相就是一个完整的步态。

GMM是多个高斯概率密度函数的线性组合。为满足奶牛步态时相识别的鲁棒性，需要奶牛长时间佩戴传感器，采集的测量数据包含静立相S_1和连续步态中的摆动相S_2和站立相S_3，总共3个步态时相，每个时相对应一个高斯分布，测量数据符合这3个时相的高斯混合分布。使用该模型的目的是通过训练找到相应参数θ，进而将高斯混合分布分簇，得到原始的每一个高斯分布，实现类似分类的效果，进而将不同的状态分类。高斯混合模型的实现过程即为参数θ的估计，采用期望最大化（Expectation Maximization，EM）算法进行参数估计[18]，该方法基于极大似然估计算法可表示为：

$$P(X|\theta) = \sum_{i=1}^{n} \omega_i \Phi_i(X) \tag{1}$$

$$\Phi_i(X) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu_i)^T \Sigma_i^{-1}(X-\mu_i)\right\} \tag{2}$$

$$\theta = \{\omega_i, \mu_i, \Sigma_i\};\ i = 1, 2, \cdots, n \tag{3}$$

其中，n是GMM的阶数，本研究中n=3；ω_i是权值，且$\omega_i > 0$，$\sum_{i=1}^{n}\omega_i = 1$；$\Phi_i(X)$是第$i$个时相的高斯分布概率密度函数；$D$、$\mu$、$\Sigma$分别是数据的维数、期望和协方差。

HMM[19]由马尔科夫模型发展而来，是一种动态贝叶斯的生成模型以及有向图模型，由隐状态序列和观测序列构成，其中每个观测数据代表一个状态。核心思想是由于状态序列的不可直接

(a)AX

(b)AY

(c)AZ

(d)GX

(e)GY

(f)GZ

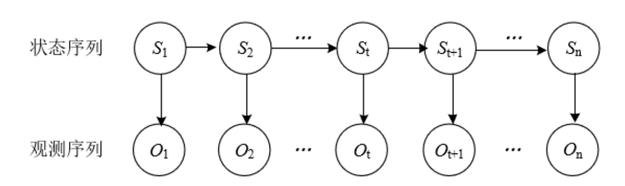
注：AX、AY、AZ分别为x、y、z轴的加速度；GX、GY、GZ分别为x、y、z轴的角速度

图4 传感器6轴数据折线图

Fig. 4 Sensor 6-axis data line chart

观测的性质，通过研究观测序列来辨识不可见隐状态序列，如图5所示。

HMM由三个参数构成：初始状态概率π、



注：S_i为隐状态，O_i为观测状态，i=1，2，3……n

图5 隐马尔科夫模型结构

Fig. 5 Hidden Markov Model structure

状态转移概率A以及发射概率（隐状态发生观测事件的概率）B。π和A构成了隐状态之间的状态转移，如图6所示。其中，S_1为静止相，S_2、S_3为连续步态中的摆动相和站立相，a_{ij}为第i个状态转移到第j个状态的概率i，j=1，2，3。B将隐状态和观测序列联系，观测序列由状态序列生成。该参数集可以表示为：

$$\lambda = \{\pi, A, B\} \tag{4}$$

HMM可以解决三个问题：

（1）评估问题。给定观测序列和参数集λ，

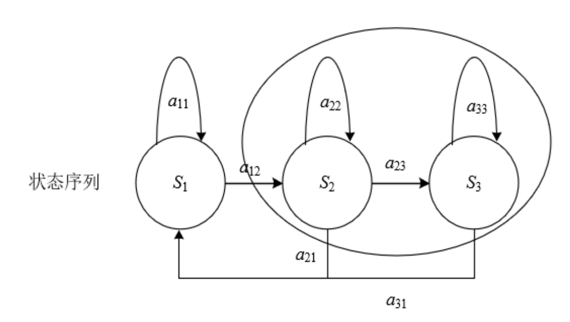


图6 隐马尔科夫模型状态转换的过程

Fig. 6 State transition processof Hidden Markov Model

计算观测序列出现的概率，采用前向-后向（Forward-Backword）算法实现。

（2）学习问题。给定观测序列，估计模型参数集λ，方法的核心是使得观测序列出现的概率最大的参数即为模型参数，采用极大似然估计的方法（EM算法或者Baum-Welth算法）实现。

（3）解码问题。已知参数集λ和观测序列，计算概率最大的状态序列，采用Viterbi算法实现。Viterbi算法本质上是一个动态规划的最短路问题，找到使得状态序列发射概率最大即由观测状态序列最可能对应的状态序列。

在本研究中，先进行学习问题的训练，利用训练集得出参数集，再基于得到的模型参数，利用测试集对观测序列进行解码。奶牛的观测数据为观测序列，奶牛的步态时相序列为状态序列，步态时相为隐含状态。为解决奶牛步态时相分割的鲁棒性，本研究针对的是长时间的连续观察序列，采用GMM与HMM融合模型（GMM-HMM），此时参数集中的发射概率B由一组观察值的高斯分布密度函数替代，即发射概率B用高斯混合模型的参数进行替代，新的参数集为：

$$\lambda = \{\pi, A, \omega, \mu, \Sigma\} \quad (5)$$

其中，π、A分别是HMM的初始概率和状态转移概率；ω、μ和Σ分别为观测序列高斯混合分布的权重，均值和协方差。

将GMM-HMM中高斯分布数设定为3，对应静立相、摆动相和站立相3种步态时相。将构建好的数据集提取特征值，将z轴角速度的1阶差分作为测试集带入模型，进行学习问题。首先对参数初始概率π和转换概率A进行初始化，然后根据EM算法反复迭代训练得出参数ω、μ和Σ，此时已经得出3种高斯分布，即3种状态，π直接按照输入的训练集初始状态定义，再根据应用统计的方法计算得出A，公式如下：

$$a_{ij} = \frac{N_{ij}}{T} \quad (6)$$

其中，a_{ij}是状态i转移到状态j的概率；N_{ij}、T分别是状态i转移到状态j的个数、各个状态之间转换的总个数。

GMM-HMM使用Python3.8中hmmlearn.hmm模块实现。

2.5 奶牛步态时相识别

经过对奶牛行走的大量观察以及查阅资料可知，奶牛的单个步态周期为1.2 s左右[20]。为满足步态时相识别的自适应性和鲁棒性，需要对连续长时间序列的步态数据进行研究。构建训练集时，考虑到有监督学习的训练方式需要人工标注的数据量过于庞大且效率低，因此本研究采用无监督学习的训练方式。通过GMM-HMM融合算法，从传感器的运动学数据中找到满足条件的特征值，对连续长时间序列数据进行无监督的聚类以及参数集训练，得到模型参数，用测试集验证结果，与标注数据对比，得出识别结果，确定奶牛步态时相，并实现奶牛步态分割。具体流程如图7所示。

3 结果与分析

3.1 特征值选择

对奶牛步态的运动学数据进行统计化处理。将2.3中选取的z轴角速度采用构建滑动时间窗和不使用构建滑动时间窗的方式，提取数据的最值、均值、方差、标准差、1阶差分、线性回归的斜率和截距，以及平均绝对误差。最终发现不加窗的情况下，1阶差分可以很好地区别奶牛步态不同时相的数学特征。图8为不同特征值的分布比较情况，截取的是连续步态周期的一段，两

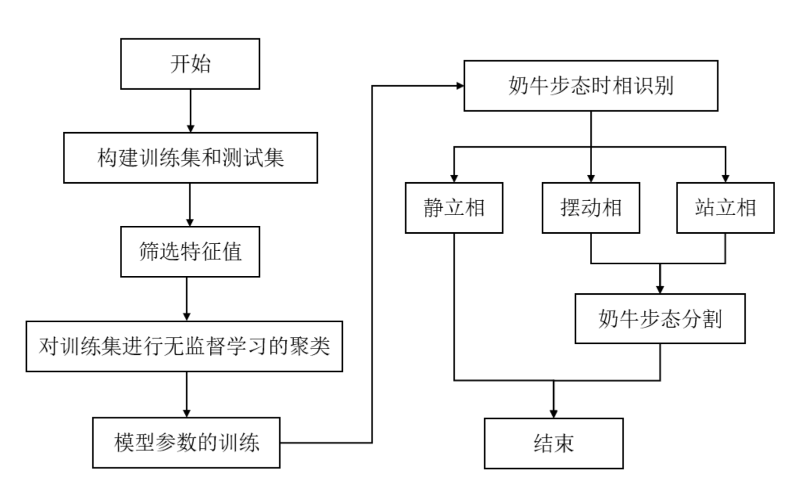


图7 奶牛步态时相的自适应识别方法流程图

Fig. 7 Flow chart of adaptive recognition method for gait phase of dairy cows

侧离散程度大的是摆动相，两条黑线之间离散程度小的是站立相。由图8可知，偏度在站立相和摆动相没有区分度，最值和均值的重合度过高且没有区分两个相位的直观特征，标准差区分度和精度不足，只有1阶差分可以直观且高精度的区分站立相和摆动相，其中站立相数据较为平缓，摆动相数据波动较大。

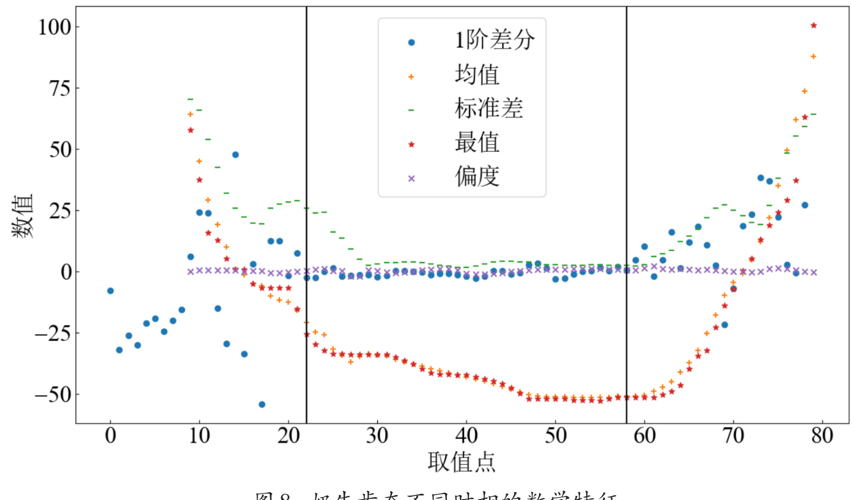


图8 奶牛步态不同时相的数学特征

Fig. 8 Mathematical characteristics of different gait phases of dairy cows

3.2 模型参数训练结果

经过训练得到的模型参数包括不同时相的初始概率、高斯混合模型的权重、均值和方差，如表1所示，公式（7）表示步态时相的转移概率矩阵 $\boldsymbol{A}$。转移概率矩阵按照静立相、站立相、摆动相的顺序排列，表示时相之间相互转换的概率。

表1 GMM-HMM模型参数

Table 1 GMM-HMM Model parameters

时相	初始概率	权重	均值	方差
静立相	0	1	0.002	0.027
站立相	0	1	−0.251	3.904
摆动相	1	1	0.357	1976.724

$$A = \begin{bmatrix} 0.987 & 0.012 & 0 \\ 0.064 & 0.904 & 0.031 \\ 0 & 0.046 & 0.953 \end{bmatrix} \quad (7)$$

从所得参数的结果可知，初始概率即为训练集的初始时相的概率，模型构建时以摆动相为初始输入时相，满足随机性。三个时相的高斯混合模型的权重都为1，说明每一阶段单独表示相应时相，没有时相之间的交叉影响，提高了模型的可解释性和准确性。转移概率的数值也表明了模型构建的准确性。每一组均值和方差代表一个时相。

3.3 奶牛步态时相识别结果

参数训练完成后进入解码阶段。将测试集带入GMM-HMM识别奶牛步态时相，得到的部分测试结果如图9所示。图中红色、绿色和蓝色部分为奶牛的静立相、站立相和摆动相。从图9可以直观发现，较为平缓的是奶牛没有行走的静立相；站立相是奶牛蹄着地到蹄离地的这一过程，其角速度与图中体现的一致，从蹄着地开始减少再增加直到蹄离地；连续步态的站立相和摆动相交替出现，摆动相的角速度增加到一个步态周期的最大值，然后减少，直到蹄落地开始受力为止。

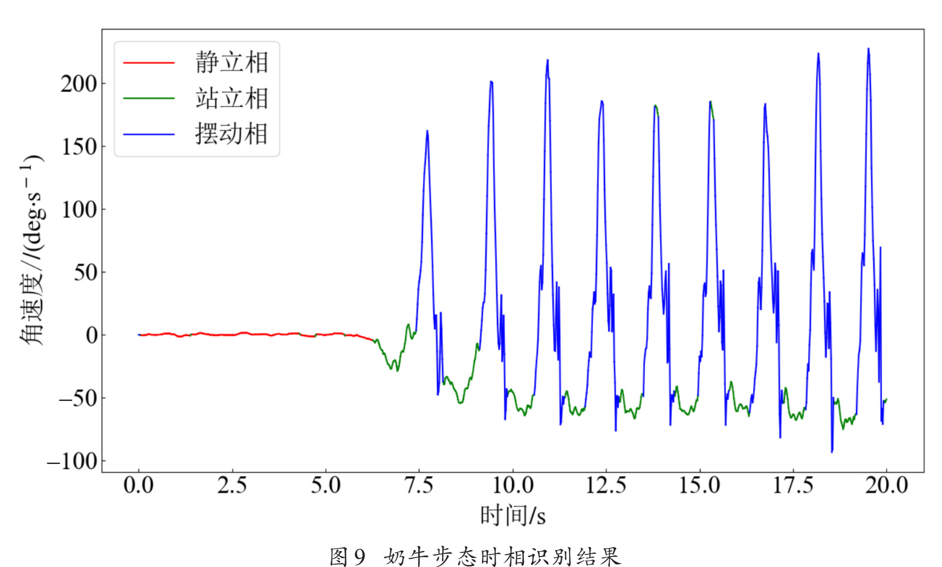


图9 奶牛步态时相识别结果

Fig. 9 Gait phase recognition results of dairy cows

采用精确率（Precision，P），召回率（Recall，R），特异率（Specificity，Sp），敏感性（Sensitivity，Se）和F_1值为评价指标。本方法的步态时相识别结果中每一条测试集数据都会得到一个状态，根据测试视频标注结果来验证测试集，结果如表2所示。

由表2结果可知，3种时相在本研究方法的F_1值相差不大，静立相的F_1值较高为90.91%，识别效果最好，说明静立相与连续步态的两个时相区分效果较好；摆动相的召回率和敏感性最高，都为91.51%，说明摆动相的特征较为明显，识别成功率最高；站立相的准确率和特异性最高，分别为91.55%和95.51%。3个相位识别结果的误差来源主要在于数据标注时数据质量以及试验环境因素，例如，视频清晰度不足以准确判断奶牛蹄离地、蹄落地这些事件引起的误差以及搜集数据时牛场湿软地面引起奶牛二次发力以及

表2 GMM-HMM的奶牛不同步态时相识别结果

Table 2 GMM-HMM recognition results of different gait phases in dairy cows

时相	P/%	R/%	Sp/%	Se/%	F_1/%
静立相	89.28	90.95	93.83	90.95	90.91
站立相	91.55	86.71	95.51	86.71	89.06
摆动相	86.67	91.51	93.58	91.51	89.02

站立不稳等因素。

通过检索文献发现，大部分针对奶牛步态的研究关注奶牛步态分割方法为提取跛行奶牛步态时相特征，本研究更进一步研究奶牛步态时相识别算法。试验通过人工引导奶牛的方式，采集了奶牛在不同行走速度的步态惯性测量数据，结果表明本研究提出的方法能够克服行走速度对步态时相识别结果的影响，准确识别步态时相，有望实现非约束环境下奶牛步态时相的自适应识别。

3.4 奶牛步态分割结果对比分析

针对奶牛连续步态阶段中的站立相和摆动相进行奶牛步态分割。在奶牛连续步态中，站立相和摆动相构成一个完整的步态周期，根据连续步态中的站立相和摆动相交替出现，奶牛步态时相识别连续步态阶段中的站立相和摆动相的结果即为奶牛步态分割的结果。奶牛步态分割时不再以每条数据为标准，使用成功分割的步态周期为单位进行结果评价，此时的标准为人工标注的完整步态周期。将成功识别的步态周期设定为正样本，连续步态中验证162步，设定误差为5帧（±0.1 s）。为区别3.3中奶牛步态时相识别的结果，重新定义*TP**为准确识别单个步态周期的个数，*FP**为非单个步态周期，但被识别为步态周期的个数，*FN**为未被成功识别的步态周期，其中*TP**为121个，*FN**为35个，*FP**为11个，此时*TN**无意义，得到准确率为91.67%。

同时，将本研究利用传感器的奶牛步态分割方式准确率结果与基于图像处理的奶牛步态分割算法，包括支持向量机、K最邻近法以及随机森林等进行对比。峰值检测法奶牛步态分割准确率为87.44%，动态时间规整奶牛步态分割准确率为90.57%[13]，本研究奶牛步态分割准确率为91.67%，不仅识别了奶牛步态时相，而且相较于峰值检测法和动态时间规整分别将奶牛步态分割准确提高4.23%和1.1%。与基于图像处理的奶牛步态分割算法支持向量机、K最近邻和随机森林的准确率分别为87.60%、83.90%和82.56%，本研究方法分别提高了4.07%、7.7%和9.11%。

本试验步态分割成功识别160步，蹄落地的总误差为-14.9 s，平均每步误差0.093 s，蹄离地的总误差为-7.2 s，平均每步误差-0.045 s，满足0.1 s的误差。其中“-“表示滞后（视频和传感器都为50 Hz）。设立的正样本是站立相，站立相是奶牛蹄着地到奶牛蹄离地这一过程，蹄着地时奶牛会出现站立不稳、滑行，以及二次用力重新站稳的现象，导致蹄着地这一过程误差较大（即站立相开始阶段），模型对蹄着地这一事件判别不准确，蹄落地后，传感器会再次收到抖动的信号，直到稳定状态，才会识别站立相，进而导致蹄着地这一事件识别会滞后。虽然蹄离地这一事件，不会有二次受力的情况，误差会大幅减小，但是松软、粘连的地面情况，也会出现较小的滞后情况。

4 结 论

本研究利用可穿戴式惯性测量单元采集奶牛后肢的3轴加速度和3轴角速度数据，经过数据预处理和特征值甄别，选择z轴角速度的1阶差分作为特征值，通过GMM-HMM融合算法进行奶牛步态时相的识别。结果为奶牛步态时相和奶牛步态分割提供了理论依据，具体结论如下。

（1）提出了一种GMM-HMM融合的无监督学习奶牛步态时相识别算法，识别出了奶牛的静立相、连续步态阶段的摆动相和站立相，其中静立相识别的准确率、召回率和F_1分别为89.28%、90.95%和90.91%，连续步态中的站立相识别的准确率、召回率和F_1分别为91.55%、86.71%和89.06%，连续步态中的摆动相识别的准确率、召回率和F_1分别为86.67%、91.51%和89.03%。进一步识别了奶牛步态时相，且识别效果较好，为后期奶牛跛行特征提取提供了技术方法。

（2）根据连续步态的站立相和摆动相实现了奶牛步态分割，准确率达到了91.67%，相较于基于事件的峰值检测法和动态时间规整算法准确率分别提高了4.23%和1.1%，分割效果较好。

为满足非约束条件下步态识别的应用要求，

今后将重点测试奶牛在不同的行走速度、跛行状态以及地面类型的条件下的模型性能，以提升模型的鲁棒性，并结合跛行引起的奶牛躺卧、采食等日常行为波动状况分析，提高跛行判别的准确性。

参考文献：

[1] TADICH N, FLOR E, GREEN L. Associations between hoof lesions and locomotion score in 1098 unsound dairy cows[J]. The Veterinary Journal, 2010, 184 (1): 60-65.

[2] OEHM A W, KNUBBEN-SCHWEIZER G, RIEGER A, et al. A systematic review and meta-analyses of risk factors associated with lameness in dairy cows[J]. BMC Veterinary Research, 2019, 15(1): ID 346.

[3] 邵军, 林为民, 孙新文, 等. 石河子地区奶牛肢蹄病类型和发病情况调查分析[J]. 中国奶牛, 2017(1): 30-33.
SHAO J, LIN W, SUN X, et al. Investigation and analysis of limb hoof disease types and incidence in dairy cows in Shihezi area[J]. China Dairy Cattle, 2017 (1) : 30-33.

[4] 韩书庆, 张晶, 程国栋, 等. 奶牛跛行自动识别技术研究现状与挑战[J]. 智慧农业(中英文), 2020, 2(3): 21-36.
HAN S, ZHANG J, CHENG G, et al. Research status and challenges of automatic identification technology for cow lameness[J]. Smart Agriculture, 2020, 2 (3) : 21-36.

[5] 程国栋, 吴建寨, 邢丽玮, 等. 基于IMU的细粒度奶牛行为判别[J]. 中国农业大学学报, 2022, 27(4): 179-186.
CHENG G, WU J, XING L, etc. The behavior discrimination of fine-grained dairy cows based on IMU[J]. Journal of China Agricultural University, 2022, 27 (4): 179-186.

[6] LI Z, ZHANG Q, LYU S, et al. Fusion of RGB, optical flow and skeleton features for the detection of lameness in dairy cows[J]. Biosystems Engineering, 2022, 218: 62-77.

[7] ALSAAOD M, NIEDERHAUSER J J, BEER G, et al. Development and validation of a novel pedometer algorithm to quantify extended characteristics of the locomotor behavior of dairy cows[J]. Journal of Dairy Science, 2015, 98(9): 6236-6242.

[8] ZHAO K, BEWLEY J M, HE D, et al. Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique[J]. Computers and Electronics in Agriculture, 2018, 148: 226-236.

[9] 宋怀波, 姜波, 吴倩, 等. 基于头颈部轮廓拟合直线斜率特征的奶牛跛行检测方法[J]. 农业工程学报, 2018, 34(15): 190-199.
SONG H, JIANG B, WU Q, et al. Detection method of cow lameness based on fitting straight line slope characteristics of head and neck contour[J]. Transactions of the CASE, 2018, 34 (15) : 190-199.

[10] JIANG B, SONG H, HE D. Lameness detection of dairy cows based on a double normal background statistical model[J]. Computers and Electronics in Agriculture, 2019, 158: 140-149.

[11] PENG Y, KONDO N, FUJIURA T, et al. Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units[J]. Computers and Electronics in Agriculture, 2019, 157: 247-253.

[12] HALADJIAN J, BRÜGGE B, NÜSKE S. An approach for early lameness detection in dairy cattle[C]// The 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and The 2017 ACM International Symposium on Wearable Computers. New York, USA: Association for Computing Machinery, 2017.

[13] 苏力德, 张永, 王健, 等. 基于改进动态时间规整算法的奶牛步态分割方法[J]. 农业机械学报, 2020, 51(7): 52-59.
SU L, ZHANG Y, WANG J, et al. Gait segmentation method for dairy cows based on improved dynamic time warping algorithm[J]. Transactions of the CSAM, 2020, 51 (7) : 52-59.

[14] 张向刚, 唐海, 付常君, 等. 一种基于隐马尔科夫模型的步态识别算法[J]. 计算机科学, 2016, 43(7): 285-289, 302.
ZHANG X, TANG H, FU C, et al. A gait recognition algorithm based on hidden Markov model[J]. Computer Science, 2016, 43 (7) : 285-289, 302.

[15] HAJI GHASSEMI N, HANNINK J, MARTINDALE C F, et al. Segmentation of gait sequences in sensor-based movement analysis: A comparison of methods in Parkinson's disease[J]. Sensors (Basel), 2018, 18(1): ID 145

[16] ROTH N, KUDERLE A, ULLRICH M, et al. Hidden Markov model based stride segmentation on unsupervised free-living gait data in Parkinson's disease patients[J]. Journal of Neuro Engineering and Rehabilitation, 2021, 18(1): ID 93.

[17] 王东源, 王玉舜, 谷禹, 等. 天津地区奶牛蹄病分析研究[J]. 天津农学院学报, 2015, 22(2): 37-39.
WANG D, WANG Y, GU Y, et al. Analysis of cow hoof disease in Tianjin[J]. Journal of Tianjin Agricultural University, 2015, 22(2): 37-39.

[18] XIA R, ZHANG Q, DENG X. Multiscale Gaussian convolution algorithm for estimate of Gaussian mixture model[J]. Communications in Statistics-Theory and Methods, 2018, 48(23): 5889-5910.

[19] RABINER L R. A Tutorial on Hidden Markov models and selected applications in speech recognition[M]// WAIBEL A, LEE K-F. Readings in speech recognition. San Francisco: Morgan Kaufmann, 1990.
[20] ALSAAOD M, LUTERNAUER M, HAUSEGGER T, et al. The cow pedogram—Analysis of gait cycle variables allows the detection of lameness and foot pathologies[J]. Journal of Dairy Science, 2017, 100(2): 1417-1426.

Gait Phase Recognition of Dairy Cows based on Gaussian Mixture Model and Hidden Markov Model

ZHANG Kai, HAN Shuqing, CHENG Guodong, WU Saisai, LIU Jifang*

(*Agricultural Information Institute of Chinese Academy of Agricultural Sciences*/ *Key Laboratory of Agricultural Blockchain Application*, *Ministry of Agriculture and Rural Affairs*, *Beijing* 100081, *China*)

Abstract: The gait phase of dairy cows is an important indicator to reflect the severity of lameness. IThe accuracy of available gait segmentation methods was not enough for lameness detection. In this study, a gait phase recognition method based on Gaussian mixture model (GMM) and hidden Markov model (HMM) was proposed and tested. Firstly, wearable inertial sensors LPMS-B2 were used to collect the acceleration and angular velocity signals of cow hind limbs. In order to remove the noise of the system and restore the real dynamic data, Kalman filter was used for data preprocessing. The first-order difference of the angular velocity of the coronal axis was selected as the eigenvalue. Secondly, to analyze the long-term continuous recorded gait sequences of dairy cows, the processed data was clustered by GMM in the unsupervised way. The clustering results were taken as the input of the HMM, and the gait phase recognition of dairy cows was realized by decoding the observed data. Finally, the cow gait was segmented into 3 phases, including the stationary phase, standing phase and swing phase. At the same time, gait segmentation was achieved according to the standing phase and swing phase. The accuracy, recall rate and F_1 of the stationary phase were 89.28%, 90.95% and 90.91%, respectively. The accuracy, recall rate and F_1 of the standing phase recognition in continuous gait were 91.55%, 86.71% and 89.06%, respectively. The accuracy, recall rate and F_1 of the swing phase recognition in continuous gait were 86.67%, 91.51% and 89.03%, respectively. The accuracy of cow gait segmentation was 91.67%, which was 4.23% and 1.1 % higher than that of the event-based peak detection method and dynamic time warping algorithm, respectively. The experimental results showed that the proposed method could overcome the influence of the cow's walking speed on gait phase recognition results, and recognize the gait phase accurately. This experiment provides a new method for the adaptive recognition of the cow gait phase in unconstrained environments. The degree of lameness of dairy cows can be judged by the gait features.

Key words: dairy cow lameness; gait phase; gait segmentation; gaussian mixture model; hidden Markov model; Kalman filtering